



AN OPEN SPECIFICATION FROM V — AUTONOMOUS INSTITUTIONS LAB

Agent Residency.

The missing accountability layer.

Identity, delegation, authorization, and audit for AI agents — anchored to responsible legal entities. Open specification. Reference implementation in build.

CORE SIGNAL

**Identity was built for humans.
Agents are not humans.**

A briefing from V on the missing accountability layer for AI agents. Authored by Vattan PS, architect of Agent Residency. Pilot inquiries open at ai@v.ee with subject line "Pilot — [domain]."

V's read of the field.

Structural, not incidental.

V's synthesis across the Cloud Security Alliance, Gravitee, Gartner, BlueRock, and 2025–2026 security disclosures. The numbers below are not anomalies. They are the steady state of an identity stack designed for humans, now asked to govern agents.

01 **18% — confidence in agent IAM**

Cloud Security Alliance survey of 285 IT and security professionals. Only 18% are highly confident their identity stack can manage agent identities. The other 82% are operating in the dark.

02 **22% — organizations treating agents as independent identities**

Gravitee 2026 report. The remaining 78% manage agents as static API keys, shared service accounts, or reused user credentials. Three structural failures of the existing stack, all in production today.

03 **36.7% — Model Context Protocol servers vulnerable**

BlueRock disclosure. Server-side request forgery exposure in 36.7% of MCP servers tested. The attack surface for agent tool use is operational, not theoretical — and the accountability layer is the precondition for fixing it.

04 **40% — agentic AI projects Gartner forecasts will be canceled**

By the end of 2027. The technical capacity exists. The identity, authorization, and audit infrastructure does not. Most projects will not survive the gap.

The three-question test.

Apply this to any vendor claiming agent identity.

Without verifiable identity, no delegation. Without delegation, no accountability. Without accountability, no trust.

Every workable agent identity layer collapses into three questions. V's diagnostic is the test. Vendors answer pieces of them as features. No vendor answers them as an accountability layer that survives investigation, regulation, or scale.

01 Which agent did this?

Today: shared API keys, reused human credentials, no per-agent identifier, no cryptographic proof. From every service the agent touches, there is no way to tell whether a human or their agent acted. Forensic dead end.

02 Who does the agent represent?

Today: the agent acts as the user. No machine-readable delegation. No bounded mandate. No revocation chain. When sub-agents spawn, the original principal is unreachable by the third hop.

03 What is it allowed to do right now?

Today: the agent inherits the human's full credentials. No scope. No expiry. No prohibitions. AI Act Article 14 oversight is mathematically impossible at machine speed without machine-readable mandates.

The agent identity stack.

What's covered.
What's not.

Six clusters, mapped against where the existing identity stack ends and where Agent Residency picks up. Incumbents extend clusters one and two. Clusters three through six are undefended — the work the existing paradigm architecturally cannot complete.

DESIGN IMPLICATION

The base assumptions were wrong. Patches inherit the assumption.

01 / IDENTITY

Identity primitives

Per-agent identifiers, cryptographic keys, lifecycle. Incumbents extend their human directories; V binds identity to a responsible legal entity.

02 / AUTHENTICATION

Authentication evidence

Verifiable proof an agent is the agent it claims to be. OAuth extensions and workload identity reach this far. The gap opens past it.

03 / AUTHORIZATION

Delegation & mandate

Machine-readable mandates: scope, caps, expiry, approvals, revocation. The piece RBAC cannot model and OAuth was never shaped for. The core of Agent Residency.

04 / AUDIT

Attribution & traceability

Tamper-evident logs of who delegated what, to whom, with what limits. Cryptographic integrity. Replayable under regulatory review.

05 / MULTI-AGENT

Delegation chains

Monotonic attenuation, propagated revocation, sub-agent accountability. The distributed-systems security problem applied to AI — and the weakest link in production today.

06 / GOVERNANCE

Threats & legal gaps

Active attacks, liability frameworks, AI Act Article 14 compliance. Where supervision and policy-as-code converge into the accountability layer.

The existing stack was built for humans.

The assumptions cannot be patched.

Epicycles on a Ptolemaic model. The base assumption is wrong, and the fixes inherit the assumption.

Agent identity features are appearing inside model platforms, IAM systems, and productivity suites. They prove the need. They do not create an accountability layer across agents, principals, mandates, revocation, and audit. The shape of the patch is wrong because the shape of the original problem is wrong.

01 **Logged-in user, single session**

Existing IAM models a directory of long-lived human identities authenticating per session. Agents may live for seconds, exist as deployments, and act across sessions in parallel. Per-session reasoning breaks at the first hop.

02 **Consent at the moment of action**

Click-to-agree, sign-with-PIN, accept-the-prompt — all assume a human reading the screen. Agents do not read screens for legal effect. AI Act Article 14 oversight is not buildable on per-action consent at machine speed.

03 **Role-based access control**

RBAC assumes static role assignments revoked through HR processes. Agents need scoped, time-bounded, revocable mandates that propagate through delegation chains with monotonic attenuation. RBAC cannot model the shape.

DESIGN PRINCIPLE

Delegation, not impersonation. Agents identifiable as agents. Bound to a responsible legal entity. Producing audit trails that survive forensic review.

One specification.

One reference implementation. Standards-aligned.

V publishes the open specification, builds the reference implementation, and maps Agent Residency to the standards regulated identity systems already trust. The model is one component. The institution is the system around it.

FROM THE ARCHITECT

The question is not whether agents are people. The question is whether systems can prove which agent acted, under whose authority, within what mandate. Today they cannot. Agent Residency is the specification that lets them.

– VATTAN PS

01 / SPECIFICATION

Open specification

IN MOTION

Identity, delegation, authorization, and audit. Versioned in public. Governed collaboratively. The open layer agents can be issued against, mandated through, and audited within.

- Mandate object: scope, caps, expiry
- Delegation graph: parent → child → revocation
- Audit schema: tamper-evident events

02 / REFERENCE IMPLEMENTATION

Reference implementation

IN BUILD

Open-source. The free, functional sample developers build on and the foundation Agency.AI's commercial product extends. Adoption driver, not revenue line.

- Mandate issuance + verification
- Per-action authorization checks
- Revocation propagation primitives

03 / STANDARDS ALIGNMENT

Standards alignment

TRACKING

Agent Residency carries existing identity infrastructure forward. No parallel stack. No vendor lock-in. The specification maps to where regulated identity is already going — including NIST's 2026 concept paper on software and AI agent identity and authorization.

- eIDAS 2.0 / EUDI Wallet
- W3C Verifiable Credentials
- NIST NCCoE / OAuth 2.0 / SPIFFE

PRIMITIVE 01

Agent identity & lifecycle

Non-human identity, keys, responsible-entity binding.

PRIMITIVE 02

Machine-readable mandates

Act-on-behalf-of with scope, caps, expiry, approvals.

PRIMITIVE 03

Authorization & enforcement

Scope evaluation, per-action checks, monotonic attenuation across chains.

PRIMITIVE 04

Audit & revocation

Tamper-evident events, fast revocation, replayable record.

What a pilot with V looks like.

One workflow per pilot. End-to-end delegation without impersonation. The agent never uses the principal's credentials. Every action is logged with cryptographic integrity. Revocation propagates. The pilot demonstrates the specification under load, inside a regulated workflow.

TRACK 01 / COMPANY MANAGEMENT

e-Resident company operations

An agent retrieves tax data from EMTA, pre-fills the annual report, and routes for the director's signature. Anything with legal effect remains human-signed.

TRACK 02 / BANKING & TREASURY

Agentic payments under supervision

Agents executing transactions and managing treasury under scoped financial authority — transaction limits, approved counterparties, audit trails satisfying Finantsinspeksioon and AML.

TRACK 03 / HEALTHCARE DATA

Clinical agents under GDPR

An agent accesses patient records on behalf of a named clinician. Strictest identity binding. Every access logged for GDPR Article 30. Unauthorized access architecturally distinguishable.

PHASE 01 · WEEKS 1-4

Architecture & workflow scoping

V designs the identity, mandate, and audit architecture for the chosen workflow. Joint review at week four locks the pilot boundary.

PHASE 02 · WEEKS 5-12

Authentication & authorization prototype

Working prototype: identity issuance, mandate enforcement, tamper-evident logging, live revocation drill. Sandbox endpoints where production isn't yet possible.

PHASE 03 · WEEKS 13-20

Pilot under load

Pilot runs against the real workflow with full audit capture. Output: evidence pack, regulatory readiness analysis, public specification contribution.

ENGAGEMENT

Spec contributor Feedback on drafts, extension proposals, public reviews of the open specification.

Spec adopter Implement the open specification on top of existing identity infrastructure.

Pilot partner Run a regulated workflow end-to-end with V. One domain, one quarter.

PILOT INQUIRY

ai@v.ee

Subject line: "Pilot — [domain]." Bring one regulated workflow. One quarter. An evidence pack on the other side.

Vattan PS

Architect, Agent Residency

V.EE · AGENTRESIDENCY.COM
LINKEDIN · GITHUB